## **Business Understanding**

### **Business Overview**

MTN Cote d’Ivoire is a leading telecom operation company (OpCo) in Ivory Coast and is part of the global MTN operations in Middle East and Africa.

MTN Cote d’Ivoire would like to upgrade its technology infrastructure for its mobile users in Ivory Coast.

### **Business Objective**

The main objective of this report is to identify the best strategy for MTN Cote d’Ivoire to upgrade infrastructure within its cities.

### **Business Success Criteria**

To compile a list of cities that will increase the return on investment of the upgrade resulting in collecting more revenue from the areas with growth portential.

### **Assessing the Situation**

1. **Resource Inventory**
   1. Datasets:
      1. cells\_geo\_description.xlsx [[Link]](https://drive.google.com/a/moringaschool.com/file/d/1-rIM5ihDu79RaH7rAs-d-7SQSAQhrY9N/view?usp=sharing)
      2. cells\_geo.csv [[Link]](https://drive.google.com/a/moringaschool.com/file/d/1ABZux280OjL3yWcOn8BDA_f5QsyO0QPU/view?usp=sharing)
      3. CDR\_description.xlsx [[Link]](https://drive.google.com/open?id=1cVoNXl25IO5-_yQk97ThdeqhE6yw8YTD)
      4. CDR 20120507 [[http://bit.ly/TelecomDataset1]](http://bit.ly/Telcom_dataset1)
      5. CDR 20120508 [[http://bit.ly/TelecomDataset2]](http://bit.ly/Telcom_dataset2)
      6. CDR 20120509 [[http://bit.ly/TelecomDataset3]](http://bit.ly/Telcom_dataset3)
   2. Software ( Github, Google Collaboratory)
2. **Assumptions**
   1. The data provided is correct and up to date. There will be some cleaning up but there will be no fundamental change of the data.
   2. There is no failures in neighboring areas that can change traffic distribution pattern on a particular site.
3. **Constraints**
   1. The available is only for 3 days. These don’t represent a complete picture
      1. Each day is not a full 24-hour cycle that will reflect true picture of subscriber bahaviour.
      2. The 3 days cannot represent a whole month usage pattern to show true normal subscriber pattern. For example, the usage on weekends, night, weekday, day (business hours) cannot be the same.

### **Data Mining Goals**

My data mining goal for this project are: -

* Which ones were the most used city for the three days?
* Which cities were the most used during business and home hours?
* Most used city for the three days?

**Data Mining Success Criteria**

Our success criteria will be measured by: -

* I target the cities that have most usage states that have the usage value sum that we will use as a priority list for upgrading the MTN infrastructure.

## **Data Understanding**

### **Data Understanding Overview**

For this project, we are using the availed dataset by the MTN Cote d’Ivoire. These datasets are: -

* The Sites/Cells distribution per city - This dataset gives the number of cells grouped into sites per city.
* Call data records (CDR) – This data for 3 days between 23:00hrs and 01:00hrs gives the actual network usage for three services; voice, sms and data.

### **Data Description**

We have two datasets available for this project. A detailed description of the datasets is provided as follows:

* **Grand Electors by State dataset -** This dataset contains the number of electors per state. It consists of two columns; ***State*** and ***GrandElectors.*** These columns outline all the states in the country and their number of electors respectively. Since there are 52 states in the United States of America, it goes to reason that the dataset has 52 entries.
* **Population by State -** This dataset, on the other hand, focuses on the population of each state. It contains the population for each of the 52 states.

### **Verifying Data Quality**

None of the two datasets had any missing values. There were also no known data errors in the datasets.

## **Data Preparation**

These are the steps followed in preparing the data

#### **Loading Data**

Loaded the datasets from the CSV and then created an SQLite database from them.

#### **Cleaning Data**

While doing data exploration, we noticed that in Grand Electors by State dataset the values in State column were in lowercase and in the other dataset the values corresponding to the same column were in uppercase. Therefore, we decided to convert the State values in Grand Electors by State to uppercase so that it corresponds with the values in the other dataset. As a result, this would make the merging of the two datasets easier.

Furthermore, we shortened the name of “District of Columbia” state to “DC”. This was recommended by the company.

#### **Merging of the Datasets**

After cleaning the data, it was time to merge the two datasets.

#### **Deriving New Attributes**

Once merging was complete, we created a new column named “Ratio”. This column was populated by dividing the grand electors in each state with the corresponding population of the said state. Afterwards, we ordered the table in descending order of the ratios.

Next, we created yet another column called “RunningSum”. In this column, we calculated the cumulative sum of the grand electors in the top states.

## **Analysis**

During our analysis, we were able to single out the following states;

1. WYOMING
2. VERMONT
3. D.C
4. ALASKA
5. NORTH DAKOTA
6. RHODE ISLAND
7. SOUTH DAKOTA
8. DELAWARE
9. MAINE
10. NEW HAMPSHIRE
11. MONTANA
12. HAWAII
13. WEST VIRGINIA
14. NEBRASKA
15. IDAHO
16. NEW MEXICO
17. NEVADA
18. KANSAS
19. ARKANSAS
20. MISSISSIPPI
21. UTAH
22. CONNECTICUT
23. IOWA
24. ALABAMA
25. SOUTH CAROLINA
26. MINNESOTA
27. KENTUCKY
28. OKLAHOMA
29. OREGON
30. WISCONSIN
31. LOUISIANA
32. WASHINGTON
33. TENNESSEE
34. MARYLAND
35. INDIANA
36. COLORADO
37. MISSOURI
38. MASSACHUSETTS
39. MICHIGAN
40. ARIZONA

The above analysis was done using SQLite. The full analysis can be found in the following notebook.[Link]

## **Recommendations**

From our analysis, we would recommend that a candidate would prioritize the above-listed states. Our main reason behind this recommendation would be that the states listed above have the highest descending ratio and also the running total for the grand electors in these states amounts to more than half of all the grand electors in the country. Therefore, prioritizing these states would not only increase the chances of winning the election but also would increase the return on investment in the entire campaign.